DOCUMENT RESUME

ED 249 286 TM 840 645

AUTHOR Mitchell, James V., Jr.
TITLE Computer-Based Test Interpretation and the Public Interest.
PUB DATE 27 Aug 84
NOTE 13p.; Paper presented at the Annual Meeting of the American Psychological Association (92nd, Toronto, Ontario, August 24-28, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Assisted Testing; Social Problems; Standards; *Testing Problems; *Test Interpretation; *Test Reviews; Test Validity
IDENTIFIERS *Computer Based Test Interpretation; Test Publishers

ABSTRACT

Computer-based test interpretation (CBTI) is discussed in terms of its potential dangers to the public interest, problems with professional review of CBTI systems, and needed policies for these systems. Several problems with CBTI systems are outlined: (1) they may be nicely packaged, but it is difficult to establish their value; (2) they do not allow the in-depth analysis of response data that a psychologist might require; (3) decision rule data are not available from the publishers; (4) the strength and statistical significance of the relationships on which these decision rules are based are often not clear or are questionable; (5) if the decision rules are modeled after the decision processes of an expert clinician, they are even more hidden than in an actuarial system; and (6) the validity of the interpretive report cannot be established. Because of the difficulties with reviewing CBTI systems, potential users cannot be guided by the advice of professionals. Policies should be developed by the profession regarding the disclosure of information about CBTI systems and the validity of their interpretive reports. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED249286

Computer-Based Test Interpretation and the Public Interest

Presented at the Division 5 APA Symposium on

The Use of Computer-Based Test Interpretations: Prospects and Problems

August 27, 1984

James V. Mitchell, Jr., Director

Buros Institute of Mental Measurements

University of Nebraska-Lincoln

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. V. Mitchell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

At the very beginning of my presentation let me state the case in the boldest and most unequivocal language I can muster: the advent of computer-based test interpretation, hereafter referred to as CBTI, presents the field of psychology with its most serious and consequential challenge of the next decade; how we react to that challenge will test our true mettle as professionals and have major impact on our credibility with the public. In my judgment the stakes are about as high as they can be.

I would like to address my topic in terms of three major issues: (1) Characterization of Dangers to the Public Interest, (2) Problems with Professional Review of CBTI Systems, and (3) Needed Policies for these Systems.

1. Characterization of Dangers to the Public Interest. Obviously computer-based test interpretation is here to stay, has grown rapidly, and is likely to burgeon further in the future; what is required is a sober assessment of its advantages and its problems and an effort to chart a course that will assure the former and limit or avoid the latter. I will outline both the advantages and the problems in my remarks but will spend more time on the problems simply because of their urgency and their importance for the development and acceptance of CBTI.

TM 810 645

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

CBTI Advantages. I can think of at least four major advantages of CBTI. First, there is much evidence that the computer can apply decision rules more accurately and consistently than the typical human, which is a decided advantage when the consequences are so important. Secondly, it seems apparent that the computer can process concurrently a greater number and variety of variables, and process them far more rapidly in all their interactive and algorithmic complexity than the human mind can typically do; the computer's capacity for multivariate analysis lends itself extremely well to the complexities of multivariate interpretation. Thirdly, for these reasons alone it would seem that the potential of CBTI might well justify the contention of some publishers that their systems could actually serve to improve clinical diagnosis and prediction beyond the constraints imposed by the less efficient processing capacities of the human mind. And fourthly and finally and almost apologetically, it would seem that advertising claims for savings in time and effort and the possibility of increased income for the clinician are not entirely without substance, regardless of our willingness or unwillingness to recognize such claims as important.

CBTI Problems. But the problems of CBTI loom very importantly as well. As I think of these problems, I cannot help remembering a rather famous statement of Oscar Buros, which began with this sentence:

> "At present, no matter how poor a test may be, if it is nicely packaged and if it promises to do all sorts of things which no test can do, the test will find many gullible buyers" (Buros, 1961, p. xxiii).

The first problem I want to cite relates closely to this statement. CBTI systems seem to be prime candidates for being "nicely packaged" and promising "all sorts of things," and what makes the situation even more serious than it is

for the conventional test is that the evidence that would confirm or contradict the promises has still another place to be hidden or buried: in the computer interpretive program itself, which is seldom available to the test user. This additional veneer has the handy potential for obfuscating things even more than they are now. Amidst the glitter of the computer revolution, the gullible are going to need more "Truth-In-Packaging" protection than they have ever had before.

A second problem is that CBTI systems really necessitate a quite different relationship between the psychologist and his or her data. Whenever I read any automated test interpretation I have exactly the same experience. I can't read a single sentence without asking, "Where did that come from?" Is it based on a scale elevation? Is it based on a configural rule; if so, which one? Is it based on a regression equation? Is it code-related, and which one? Does it reflect a simple factor loading? What are the decision rules behind that statement or group of statements? Is the statement a simple extension of a score label? Or is it a test item itself or a paraphrase of that item? The questions are endless. Psychologists who use these reports are not doing their job if they don't ask these questions; neither are they doing their jobs if they do ask these questions, however, because it would take so long to get the answers that they wouldn't have time to do anything else! The relationship of one's judgments to one's data is an extremely important one, and the CBTI setting establishes a major difference in that relationship. As researchers we know the importance of getting deeply enmeshed in one's data to achieve any real understanding of it. I am not a clinician, but in some of my own work with people my best insights have come from analyzing the actual items that contributed to an individual's score, and sometimes encouraging the test-taker to do the same. But that kind of opportunity is long gone with most CBTI systems. In many cases you have to

accept on faith, literally, and hope and pray that the interpretive program author has contributed some parallel good works. In my judgment most of the CBTI systems put the psychologist-user in an extremely difficult position with respect to the basic data that constitute the important tools of the profession.

A third major problem is that even if the psychologist wanted to get the answers to these questions, in many or perhaps most cases this kind of decision rule data would not be available. Publishers vary in their willingness to disclose this information, but in all too many cases it is either unavailable or not in a form that would be at all comprehensible or useful to the user; mostly, I would judge, it is simply unavailable. Reviewers for the upcoming Ninth Mental Measurements Yearbook complain about this all the time. One reviewer, discussing the computer narrative report for a new personality inventory, had this to say: "The manual includes virtually no information on the development of the narrative report statements, except to say that they are largely based on item content, but also reflect clinical impressions and interpretations based on test results and interviews with officer candidates." That should not satisfy even the most naive. Another reviewer, discussing the computer narrative report of another new personality inventory, stated it this way (publisher identification deleted):

> "It is impossible to evaluate the procedures used for scoring this instrument since no data are provided by the manual which detail precisely how scoring is accomplished. One must simply trust [the publisher] to machine score and interpret the instrument without being able to review the statistical procedures and decision rules used by the computer for making various determinations."

Is it any wonder that at a recent APA-sponsored meeting for test publishers, the head of one CBTI firm reported that by far the greatest single concern expressed by his clients was not having access to the decision theories or decision rules by which the narrative report generated its statements?

The most grievous kind of situation occurs when the serious empirical weaknesses of a test are effectively buried under the claptrap of an overblown computer report. One of our reviewers cast this problem in bold relief when he made this concluding judgment: "It is ironic that an inventory that has a weak empirical foundation is available only through a computer automated report."

There are some very bright and creative people developing these CBTI systems; just as we would expect from our general knowledge about tests, these systems range from the good to the very poor. The basic problem is that without access to more decision rule data, we are helpless in our efforts to distinguish the former from the latter.

A fourth major problem with CBTI systems is that which often arises with systems that are actuarial-based in nature. Even if one knows something about the decision rules of an actuarial system, the strength and statistical significance of the relationships on which these decision rules are based are often not clear or are questionable or both. One not uncommonly encounters verbal statements by CBTI system developers that seem to extend rather far beyond reasonable inference from the data at hand. In a field where it takes a meta-analysis to establish even the simplest of generalizations, one sometimes wonders at their temerity. Especially when that temerity might cost someone a job. Decision rules in actuarial systems are no better than the research data or other empirical support that underlie them and provide their single claim to legitimacy.

A fifth problem with CBTI systems is that which arises with systems that are modeled after the decision processes of an expert clinician. Here we have a threefold issue: the accuracy of the modeling process; the inferred relationships between variables in the clinician's thought processes, which are even less accessible to empirical verification than in an actuarial system; and the degree of correspondence between the assumptive structures of the expert clinician and the user of the system. The latter is especially intriguing, since we seem to speak so many different languages in psychology. I hadn't noticed that we communicated too effectively across these different theoretical systems, and the spectre of miscommunication and misinterpretation as a result of divergent conceptual systems of the expert clinician and the report user is something that should receive greater attention than it has.

A sixth problem with CBTI systems is how to establish the validity of a computer interpretive report. Establishing validity is one of the holy sacraments of our profession, and yet the application of validity considerations to the computer interpretive report is neither easy to achieve consensus about nor easy to do. Some take the point of view that application of traditional validity notions is simply inappropriate. In an article in the Ontario Psychologist Paitich (1970) urged his readers to "acknowledge that the traditional psychological report is never scrutinized carefully for purposes of assessing validity (although we cannot be so lax about the tests themselves). It is found either useful or not useful. The same criterion should be applied to the automated report." Here we have a new double standard. Tests should be validated, but when test results get buried in a computer interpretive report, we have no such obligation for the total product. The computer now has a new function: professional absolution or washing away our sins with computer holy water.

If we don't take the Paitich point of view, we face a very difficult task in our effort to establish interpretive report validity. Attempts to establish report validity with ratings of appropriateness and usefulness by clinicians, or with comparisons between clinical and automated interpretations, are seldom compelling or satisfying. But if we are purists about the issue and really want to do the job properly, with some CBTI systems we are faced with the task of conducting a statement-by-statement validation involving statements generated by decision rules and decision trees of almost incomprehensible complexity. Is it realistic to think that we can do this? What happens if it is not? If utility replaces validity as the criterion, what are the implications for helper-helpee relationships in the so-called "helping professions"? Are we helping people to gain new and accurate insights into themselves or engaging in fortune-telling by computer?

The dangers to the public interest in the six problems described previously are clear. These CBTI systems will touch people's lives in important ways. If there is not "goodness of fit" between the individual we seek to help and the practical results of the tools by which we hope to help that individual, it is that individual client who will bear the consequences, whatever they may be. We owe it to the public interest to capitalize as much as we can on the potential of CBTI and yet be sufficiently alert to its present shortcomings that we can work intelligently to overcome or ameliorate them.

2. Problems With Professional Review of CBTI Systems. One way in which we can protect the public interest, of course, is to make sure that we offer good professional reviews of CBTI systems for the professionals who will use them and the clients who will be affected by them. That is one of the major tasks of the Buros Institute. But in trying to accomplish that task the Institute has

encountered some problems that it must work creatively and hard to resolve. Let me now share some of those problems with you.

One of our problems with CBTI review is one we share with users of these systems: the unavailability, or practical unavailability, of the decision rules that professional reviewers need to do an adequate job of reviewing a CBTI system. In a review on "Personality Assessment" in the Annual Review of Psychology, Lanyon (1984, p. 690) reports that he made telephone calls to several of the CBTI companies to obtain information about their interpretive programs, and that these calls "were met with self-serving statements or papers that did not directly address validity, but did address any question to which some positive answer could be given." He further reported that the available literature is of three kinds: "(a) glossy promotional literature, sometimes masquerading as scientific data, and usually accompanied by sample reports; (b) studies of customer and user satisfaction, which has never been much of a problem; and (c) an occasional paper giving actual information about the development or validation of an automated system." At the Buros Institute we have had a very similar experience. As Director of the Institute I sent a personal letter to all the CBTI companies for which the Institute had addresses and asked them to explore with me ways in which we could review their systems for the Ninth Mental Measurements Yearbook. In the letter I specifically mentioned the importance of our professional reviewers having access to the decision rules governing the interpretive statements. There was a very wide variety of responses to this letter, ranging all the way from complete cooperation to ignoring the letter to indulging in various forms of gamesmanship. One company was willing to reveal the decision rules but discovered that one of its interpretive program authors was not. All in all, my subjective impression is that I encoun-

tered, on the average, noticeably more resistance than I typically do for tests. I believe that this is an ill omen and definitely not in the public interest. In the Lanyon review cited above, the author ends his discussion of automated interpretation with the statement that "Perhaps the time has now come when federal regulations for this industry are necessary for consumer protection" (Lanyon, 1984, p. 690). As a concluding statement that's a real shocker, but we may face such circumstances, and probably should face them, if we cannot put our own house in order.

At the Buros Institute we face many other problems in our efforts to review these CBTI systems. Some systems are hardware-software combinations, and the hardware obviously can't be shipped to us in the form of the three complimentary copies we usually ask for. Yet if we ask present owners to review these systems, we run into all kinds of cognitive dissonance and conflict of interest problems. Also, for the more comprehensive and complex systems, the amount of time, effort, and expense required to do an objective, thorough, and fair review could be downright prohibitive. Some people have suggested review teams, but that makes the process even more cumbersome. Another problem that we encounter far more often than we should like is that often the few people who are really knowledgeable about CBTI system development have conflict of interest problems in that they serve as consultants for competing systems. Nevertheless, our coverage of CBTI systems will probably be expanded in the 9th MMY, and we are devoting considerable thought to how we can surmount the remaining difficulties. If you have any thoughts, please let us know.

3. Needed Policies for CBTI Systems. Computer-based test interpretation is like the human being: it has great potential for both good and evil. It is here to stay; we'd better pay attention to the potential evils of CBTI and get

things in hand quickly so that we are in a better position to realize the good. What policies are needed to do this? The latest draft of the new Joint Technical Standards for Educational and Psychological Testing speaks to CBTI, but I sometimes wonder if it speaks strongly enough. The strongest statement is probably Standard 5.11, which indicates that CBTI services should provide a test rationale and a summary of evidence supporting interpretations, including data on the validity of cut scores or configural rules. Naturally I worry about the use of the word "summary" in this standard, since professional reviewers of CBTI systems are going to need rather detailed evidence of these decision rules to conduct thorough and fair reviews of them; the potential user may also need more than a "summary" just to make an informed choice.

I feel more secure about a document entitled Guidelines for Computer Based Assessment and Interpretation, which was developed under the joint auspices of the APA Committee on Professional Standards (COPS) and the Committee on Psychological Tests and Assessment (CPTA). This document calls for complete disclosure of interpretive statement libraries for professional reviewers, which I strongly urge you to support. It is not quite as emphatic on the needs of users, calling at various times for a "manual reporting the rationale and evidence in support of computer-based interpretation of test scores" and then later indicating that the manual "should provide at least a general link between the scores and interpretive statements." The "general link" terminology is not sufficiently precise to keep us out of trouble.

What is needed in fact is a well-differentiated system of information dissemination for every CBTI system that provides information at three or four different levels: Level 1 - Complete disclosure for professional reviewers; Level 2 - Basic and coherent information for the CBTI user, which provides

rationale but also provides something more than just a general notion of how interpretive statements are generated through decision rules; perhaps there should be two sublevels here, with appendix-like information for the person who wants to delve more deeply; Level 3 - Information for the test-taker, who needs to be removed from the realm of computer myth and mystery and obtain a general perspective of how the system works and its fallibilities as well as its merits. I believe that we should insist on this differentiated three- or four-level system of CBTI information dissemination, should further elaborate on it, and should incorporate such elaboration into our CBTI standards. We must protect the CBTI publisher against copyright infringements with all the vigor at our command, but we must also recognize that public interest must take precedence over proprietary interest or the matter will be taken out of our hands.

I also believe that we must take steps to appoint a committee, perhaps related to the joint COPS-CPTA committee, that will carefully study the issue of the validity of computer interpretive reports and (1) make some recommendations about what can reasonably be expected in this area, (2) generate some explicit and fairly specific standards that CBTI systems will be expected to meet, and (3) provide some methodological suggestions for how to meet these standards. If it is true, as Lanyon (1984, p. 690) says, that "The overwhelming problem is lack of demonstrated validity for the printed interpretations" and that "lack of demonstrated program validity has now become the norm," we need to provide some immediate guidance to CBTI developers. If ever a proactive stance was needed, it is needed now at this critical juncture in the development of this important tool.

References

Buros, O.K. (1961). Tests in Print. Highland Park, New Jersey: Gryphon Press.

Paitich, D. (1970). Refining computerized test interpretations: An in-depth approach. Journal of Personality Assessment, 41, 157-159.

Lanyon, R. I. (1984). Personality assessment. In M. R. Rosenzweig and L. W. Porter (Eds.), Annual Review of Psychology, 35, 667-701.